DOCUMENT RESUME

ED 464 919 TM 033 849

AUTHOR	Mintrop, Heinrich
TITLE	The Limits of Sanctions in So Called Failing Schools: A Study of Maryland and Kentucky Schools on Probation.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	2002-04-00
NOTE	44p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
CONTRACT	R308F70035
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Probation; Case Studies; Elementary Education; *Low Achievement; Low Income Groups; Middle Schools; Minority Groups; *School Restructuring; *State Action; Surveys; Teaching Methods
IDENTIFIERS	*Kentucky; *Maryland; State Takeovers

ABSTRACT

This study investigated the effect of probation on individual performance motivation, organizational processes, and patterns of instruction in schools that were on probation for low achievement in Maryland and Kentucky. Findings are based on case studies of 11 schools on probation, 7 in Maryland and 4 in Kentucky, and all 11 schools had high proportions of students from poverty and minority backgrounds. Each case study consisted of quantitative and qualitative data: interviews, classroom observations, meeting observations, and survey questionnaires. About half the schools were elementary schools; the rest were middle schools. Studies of individual learning and organizational development show that probation in the context of the 11 schools provided unfavorable conditions for learning new and ambitious performance-based pedagogy. For many teachers, the state assessments did not provide meaningful tools for the self-evaluation of their teaching. On the organizational level, probation fostered rigidity and compliance with external obligations to the detriment of organizational learning and internal dialog. While large numbers of teachers in the 11 schools viewed themselves as highly competent professionals, 70% to 80% of the observed lessons in Maryland did not show evidence of elaborate level teaching at all. On the positive side, findings show that almost all of the 11 schools were modestly energized by the probation label, and teachers in all schools reported that they increased work effort and engagement in school improvement. When asked to select priorities for school improvement, teachers cited several factors, but not even 10% believed that a new pedagogy should be the first thing on the agenda. (Contains 55 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 919

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

H. Mintrop

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE LIMITS OF SANCTIONS IN SO CALLED FAILING SCHOOLS

A Study of Maryland and Kentucky Schools on Probation

by

Heinrich Mintrop, UCLA

Paper presented at the Annual Conference of the American Educational Research Association,

New Orleans, Spring 2002

TM033849

BEST COPY AVAILABLE

The proliferation of high-stakes accountability systems in the United States has fast created a new category of schools identified with various labels: Schools on Probation, Schools under Reconstitution, Schools in Decline or in Crisis, Schools under Review, Immediate Intervention Schools, Schools Eligible for Assistance, and so on. Each accountability system has created its own nomenclature, but the underlying gestalt is the same: Based on a small set of numerical performance indicators, accountability systems identify putative underperformers that are given a limited period of time to reverse growth deficits or decline and that are threatened with more severe penalties upon failure to do so. In the public debate, these schools are also known summarily as "failing."

As of the year 2001, 27 states have had school accountability systems that identify low-performing schools and 14 states stipulate more severe penalties when an underperforming school fails to improve (Boser, 2001). In 1997 the city of Chicago alone identified a hundred or so public schools on probation that managed to have fewer than fifteen percent of their students read at the national norm, as measured by the Iowa Test of Basic Skills (Chicago Public Schools, 1997). To date, the small state of Maryland has identified 200 or so schools statewide. Between 1999 and 2001 alone, the large state of California identified 1290 persistently low-performing schools that are enrolled in the state's "Immediate Intervention/ Underperforming Schools Program." Implementation of the new federal Title I legislation may add further impetus to the phenomenon. What is more, these schools are not randomly or evenly distributed across the states, but in many instances are clustered in districts that traditionally serve poor and disadvantaged minority populations. For example, in Maryland, almost all identified schools are located in two districts; in California, 54 of the 1000 or so school districts with more than 10

schools have at least a third of their schools eligible for California's Immediate Intervention/ Underperforming Schools Program.

Leaving aside the possibility of a mere symbolic purpose of "high stakes," these policies bank on the motivational power of sanctions, currently conceptualized in two ways. In one version, popular among economists, high-stakes accountability is resource-neutral; that is, improvements occur as a result of changed orientations and dispositions towards work effort (Hanushek, 1994). In another version, strongly advocated by researchers around the Consortium for Policy Research in Education (CPRE), sanctions, such as probation or reconstitution, attain their motivational power in conjunction with resources needed for capacity building in schools that fail as much for lack of will as for insufficient capacity (Fuhrman & Elmore, 2001; O'Day, Goertz, & Floden, 1995; O'Day & Smith, 1993). Thus, in the first version, clear performance goals, incentives, and sanctions make new resources unnecessary while in the second version they make new resources more effective. But in either case, the motivational power of incentives and sanctions on individuals or organizations is assumed or implied. In fact, we could probably dispense with the whole superstructure of "high stakes" that many states have built up in the last few years and return to more traditional redistributive grant making patterns (Peterson, Rabe, & Wong, 1991) if it was not for the belief in the power of incentives and sanctions for the improvement of low-performing or "failing" schools. The power of incentives and sanctions is even more crucial in systems that place the accountability burden on schools rather than districts. In such systems, incentives and sanctions must compensate for the states' limited capacity to directly regulate or administrate remote school actors.

"Naming and shaming," as the English say, threatening more severe penalties, signaling public urgency and support are major mechanisms of probation that are to impel individual

educators and schools to improve. Generally speaking, while there is some research on the effect of high stakes accountability on schools generally (Firestone, Mayrowetz, & Fairman, 1998; Kelley, 1999; Kelley, Conley, & Kimball, 2000; Kelley & Protsik, 1997; Newmann, King, & Rigdon, 1997; Fuhrman & Odden, 2001), little research on the role of sanctions in low-performing schools is available (Hess, 1999; Hess, 2000; Ofsted, 1997; Reynolds, 1996; Wong, Anagnostopoulos, Rutledge, Lynn, & Dreeben, 1999; Wong, Anagnostopoulos, & Rutledge, 1998), despite the proliferation of the phenomenon. But the absence of research does not necessarily mean that "little or nothing is established," (Wilcox & Gray, 1996, p.3) or as researchers have it, that nothing "is known." While certainly holding little appeal to the education profession that is subjected to it, probation and sanctions must make intuitive sense for those that decree and design accountability systems. This paper, trying to get behind these intuitions, reports on findings from a three-year study of schools on probation in the states of Maryland and Kentucky. The study investigated the effect of probation on individual performance motivation, organizational processes, and patterns of instruction. The primary purpose of this paper is to provide a summary of findings in abbreviated form.[1]

The Two States

High-stakes accountability has been a topic of vigorous debate and discussion among educators and educational researchers in recent years. Particularly the Texas case has found wide attention (McNeil, 2000; Sklar, Scheurich, & Johnson, 2000). The states of Maryland and Kentucky, by contrast, garnered national acclaim (Quality Counts 2001, 2001) for centering their

---

[1] See the full report at www.gseis.ucla./faculty/mintrop/ , "Schools on Probation".

accountability systems on tests that went beyond basic literacy and numeracy by asking students to perform complex learning, experiments, cooperative projects, complex essays, and portfolios. Although both states have by now abandoned the complex tests with which they started out, our data were collected at a time when they were still in use, though as in the case of Kentucky already contested. Thus, this study informs on schools' responses to probation in pedagogically complex accountability systems.

Naturally, there is more to an accountability system than student learning assessments. There are non-academic performance indicators (in the case of the two states mainly attendance for elementary and middle schools), there are rewards and sanctions, selection criteria for low-performing schools, exit criteria for probation, school governance requirements, planning mandates, monitoring systems, and supports for building capacity at schools. These characteristics are embedded in authority relationships between schools, districts and the state. And all of these elements are in constant flux as political coalitions shift and new plans are advanced by state policy makers (Cibulka & Lindle, 2001), making accountability systems truly moving targets of study.

At the time the study was conducted between 1997 and 2000, both states had the main features of elaborate accountability systems in common: complex student assessments, performance categories for schools, rewards and sanctions as well as school improvement planning and monitoring. But within this basic structure they differed in some respects. Compared to Maryland, the Kentucky reform was more comprehensive, more rule-bound and scripted, but also more contested and in transition (Pankratz & Petrosko, 2000). The Maryland accountability system with the MSPAP (Maryland School Performance Assessment Program) as

its heart piece was more radical in its performance demands and also more consensual at the time.

Kentucky schools on probation, or in the state's language at the time "schools in decline," were identified through a straightforward formula calculated on the basis of quantitative growth expectations, and they exited the status when they met expected test score gains. Schools on all levels of absolute performance could be in decline when they did not meet their targets. Schools could attain modest, but not trivial monetary rewards for raising test scores. The state dispatched to schools "in decline" a trained change agent (called Distinguished or Highly Skilled Educator) who provided know-how on the system's requirements and mechanisms as well as general skills in school improvement (David, Kannapel, & McDiarmid., 2000). Accountability brought to Kentucky schools more managerial autonomy from districts and a new school-internal governance structure of shared decision making and parental involvement. At the same time, the state gradually increased the prescriptiveness of the state curriculum. Towards the end of the study, the format of the central test became more traditional and sanctions lost some of their rigor, having never been fully applied anyway. The reconfiguration of the system wiped schools' slates clean again.

By contrast, Maryland left wide discretion to the state department of education in selecting "reconstitution-eligible schools" (the state's term for probation), imposing sanctions, and exiting schools from probation. The state department tended to select rock bottom performers for probation, applied final sanctions very sparingly, and set exit criteria (performance at state average) very high. State rewards and supports played a lesser role in the system. During the study period, reconstitution was a tool of the state to influence the reform disposition of two large districts that were expected to provide local resources and support to

failing schools in order to avert the threat of state take-over of schools. Thus, the state exerted indirect pressure on districts to take the test seriously, but beyond that it generally provided little pedagogical guidance and capacity building. Probation evolved into a situation in which some 200 hundred schools have been languishing for years.

The Study

Findings are based on case studies of eleven schools on probation in the two states. Each case study set consists of quantitative and qualitative data: interviews, classroom observations, meeting observations, and survey questionnaires. To gain a better understanding of the behavior of a larger number of schools on probation, we also analyzed school improvement plans from 46 schools in Maryland and 32 from Kentucky as well as state test score data from Maryland. (We did not conduct such analysis for Kentucky due to the change of test formats during our study). Data collection took place between the spring of 1998 and the spring of 2000. The study investigates the role of probation in schools that serve student populations with high proportions of children from poverty and minority backgrounds. Thus, all eleven schools have high proportions of students in the Free or Reduced Lunch program. In the Maryland schools, more than 90 percent of the students are from an ethnic minority background. In the Kentucky schools, minority proportions are above the state average.

The seven Maryland and four Kentucky schools were selected according to district, school type, duration in the program, educational load, and performance history. In each state, about half of the selected schools are middle schools, half are elementary schools. In Maryland, the schools are in the two districts where almost all schools on probation are located. In Kentucky the schools reflect the state's geographic diversity. Four schools, two in each state, are

probation veterans while seven schools had been identified half a year prior to data collection. We did not select schools based on their previous performance. Rather, we wanted to study the unfolding of probation, not knowing whether the schools would be successful in their improvement effort. As a result, this is not a design that allows us to evaluate the programs in the two states.

Each school was visited numerous times by at least two researchers over a two-year period from 1998 to 2000. The database for each Maryland case typically consists of a survey, a minimum of twenty-one formal, semi-structured interviews, and many more informal ones as well as six classroom observations per school. Interviewees were teachers of all subjects, administrators, instructional specialists, and other resource teachers. All principals were interviewed. We also interviewed district officials who were responsible for programs in schools on probation, as well as state officials, state monitors, and district support personnel. At least four meetings at each school were formally observed. In many cases the researchers participated in a number of additional meetings. In the Kentucky schools we interviewed slightly fewer numbers of teachers and observed fewer classrooms. Interviews were conducted with the help of standardizing protocols and transcribed and coded with the help of NUDIST. To better understand instructional patterns we analyze data from 45 (30 MD, 15 KY) classroom visits, consisting of a lesson observation and subsequent debriefing interview.

The teacher questionnaire, containing 250 items, was administered to all full time teachers at the eleven schools. Findings from the survey data stand together with qualitative data from interviews, meetings, and classroom observations. Overall response rate to the survey was 53 percent, though response rates varied by school. Across the two states and eleven schools, a total of 287 respondents returned valid questionnaires. An analysis of respondents'

characteristics show that teachers with leadership roles in their schools are over-represented in the sample.

Individual Work Motivation

Intuitions

Although I have never seen it explicitly spelled out by policy makers and designers of accountability systems why probation and the threat of sanctions would be effective motivators in educational settings, one could imagine the following intuitive scenario: When a school is publicly labeled as deficient, teachers after going through a whole range of emotions accept the urgency of improvement. This urgency is reinforced by the discomfort caused by state audits and the like. Teachers and administrators want to repair their public image, but they also take responsibility for the quality of their work. So, they take a critical look at their own work and reflect on the valid performance demands of the accountability system. They finally decide to increase effort in their own classroom and get involved in the improvement of their school. Teachers who are highly committed to their school are especially motivated. Additional support that might accompany probation is appreciated and put to good use, but fresh resources are not essential for increasing one's effort in the classroom.

Theories

In the literature, work motivation in accountability systems is often conceptualized in relationship to sense of efficacy (Ashton & Webb, 1986), control (Weiner, 1986), goal setting (Locke, 1968), and expectancy of rewards (Lawler, 1973). These varied, though related sources of motivation (Rowan, Chiang, & Miller, 1997) are assumed to increase teachers' performance if teachers believe that the task is in their control and they have the requisite competence for its

execution, if they see a connection between individual effort and expected reward and if they value the reward itself. Teachers strive for goals that are clear, specific, worthwhile, and attainable (Kelley & Protsik, 1997), and accountability systems streamline the work situation in this regard. Shamir, on the other hand, doubts the applicability of "point of action" theories of motivation, as he calls them. These models of motivation are useful in predicting discrete task behavior, but they are less powerful in explaining a "diffuse and open-ended concept of commitment" (Shamir, 1991, p. 408) that refers to a "shifting number and range of rather ill-delineated performances rather than to ironclad and numerically constant behaviors having clearly defined parameters that everyone knows" (ibid.).

In Shamir's view, expectancy and goal setting models of motivation presuppose "strong situations," i.e. situations structured by clear and specific goals, reward expectancies, and clearly identifiable relationships between increased effort, performance, and reward. Schools, however, are "weak" performance situations in which moral purpose and internalized standards are primary motivators. If he is right, then the accountability system would become motivating to the degree that it reinforces educational goals valued by teachers. But imbued in probation is not only an incentive to improve and attain rewards, either for the sake of the children or one's own professional prestige, but also an element of coercion and a threat of further penalties to which minimum compliance or exit might be the answer (Katz, 1970; Porter, Lawler, & Hackman, 1975; Vroom, 1964). While the research has not found a clear relationship between job satisfaction and work motivation, job satisfaction is related to job commitment, i.e. one's willingness to show up for work or stay on the job (Lawler, 1973; Mohrman, Mohrman, & Odden, 1996). According to Shamir, more congruence between work motivation and

commitment may be expected if accountability systems tap into teachers' more deep-seated values, ideals, and performance standards.

Findings

Awareness. Probation had the attention of the majority of educators at the eleven schools, but especially in the long-term probationary schools knowledge of what the status entailed became sketchy. When the school's status was first announced in public, many teachers felt "very demoralized," "really down" or "mortified." Senior teachers who described themselves as hard-working were shocked: "I took it very personally, because of the efforts that I've made in the years that I've been here... It was almost like, I had broken an arm, and I was in a lot of pain that particular day" (A-19; eighth grade science teacher). But soon thereafter, personal distancing ensued and personal culpability was rejected: "I viewed it as a very negative cast over the school and over me, because I thought it was basically speaking about my instructional leadership. But then on reflection, I realized that it wasn't about me. ....So, once I cleared my head of any guilt feelings, then I was able to move forward" (E-7; elementary school principal).

Mild pressure. For most interviewees, the imposition of final sanctions was inconceivable and they were not worried about their jobs. Instead, many stressed their professional worth in spite of public perceptions. On the survey, we asked teachers to rate themselves as professionals. Overwhelming majorities said they were adequately or very well prepared (90 percent), highly skilled (60 to 70 percent), and very willing to exert effort (about 70 percent), despite the fact that large numbers of teachers particularly in the Maryland schools were fairly new to the profession (46 percent five years or less) and the school (71 percent five years or less). On the other hand, our sample is biased towards activist teachers that fulfill some leadership functions at their

schools (43 percent of sample). In the interviews, teachers voiced confidence and contended that they did the best they could under the difficult circumstances, such as this elementary school teacher in one of the first Maryland schools on probation: " Basically, if you think you can do it better, come in, step in, and feel free to show us how to do it any better than how we've been trying to do it.... They [the state] lay these threats on the table, 'We're gonna come take you over.'... And you just get to the point where you say...'Fine, fire me!'"

Most teachers perceived probation as mild pressure and did not worry about threats. Instead, for them probation signaled the need for support, and they were willing to endure the stigma in return for new resources: "The stigma is the minus, but the programs that come about from that is a plus. You know, it's kind of two-sided.... I think, you know, the programs that would come about because of it, you know, it outweighs the negative. I think it's good, but I think that they should get rid of the bad stigma that goes with it" (B-5). For most interviewees, probation was not occasion for self-searching. Rather it was a nuisance and stood as well for vague hopes for support.

We inquired in both survey and interviews about the accountability system in general. After all, it is this system that spells out the rules for rewards and punishments. We operationalized the theoretical models by asking how teachers perceived the importance, validity, fairness, realism, and directiveness of the system.[2] Roughly speaking, the goals of the accountability system are of medium to high importance for respondents. When asked in the interviews what makes attaining high test scores important, most teachers responded that it was

---

[2] These were direct questions in the interviews and scales derived from questionnaire items. (See Technical Report for details at www.gseis.ucla.edu/faculty/mintrop/).

"a prestige thing." They didn't like "being at a school where every day in the paper they say we're failures" (20-09; Kentucky middle school English), and "all the county sees is the test scores" (B-12, Maryland middle school Health). Fewer interviewees also saw the tests as a useful gauge of performance and some of them, particularly administrators, said they evaluated their success based on the scores, but for many more, the importance of the system's performance goals was connected to their concerns for diminished professional status.

Fairness. The accountability system was less connected to the quality of teaching and learning because large majorities of respondents doubted the system's fairness and validity as a gauge of good teaching. The system was seen as unfair because it did not reflect that "honestly.....it's not the teaching as much as it is the children" (10-14; first grade), and "we are doing our part, ... really, the biggest part is missing for a lot of them, that's the home" (A-11; seventh grade). Many teachers rejected the burden that accountability systems singularly placed on them and called for more distributed responsibility for student achievement: "The adage that it takes a village to raise a child is true, you know, and what that accountability thing says to me is you only hold a few villagers, instead of the entire village for the accountability" (E-8, Maryland elementary school).

Validity. Only about a quarter of the survey respondents agreed that the state assessments validly reflected "good teaching." Whereas accountability systems are designed around standardized outcomes, teachers in the eleven schools were avowedly child-centered in their philosophy. "The state, I really, the test, I could care less about, to be honest with you" (B-7). For them, "just getting these kids to do their best and be able to write and to answer questions [was their] ...key priority" (B-7). In this vein, 71 percent of respondents asserted in the survey that "rather than expecting a great improvement in school performance test scores, [they]

concentrate on individual students' growth, no matter how small." When asked to rank-order a catalog of 11 quality indicators for their work, respondents gave standardized tests ranks 7 to 9. Teachers drew their sense of success primarily from direct interaction with students, comments from parents, and teacher-made tests. Teaching "life lessons," basic skills, citizenship, and social conduct, in the eyes of interviewees essential for their clientele, were not captured by the test, but took center stage in their classrooms. Therefore "looking at the kids' background, and looking at what is written in that test and how it addresses them and the social issues that they have, they may not make that connection. So, they may not do well. But what is important to me is if my kids are learning the things that I'm teaching them, somehow they're able to connect it to the things that they're doing" (A-8; eighth grade social studies).

Thus, personalization prevailed over data-drivenness, incrementalism over the ambition of vast test-score gains, and basic skills orientation over performance-based pedagogy. Teachers' internal performance standards were not congruent with the external standards of the accountability agency. Many teachers' self-concept eschewed the image of the score maximizer in favor of the image of an educator beholden to the intellectual and social growth of individual students and committed to the needs of the local community. Likewise, rewards were derived from encounters with individual students or learning groups and from psychic satisfaction:

> I don't feel like I need to know that they think that I'm doing the best at this and they're going to reward me for this or whatever. That's just not really important to me. I like to see my students succeed and I like to think that yes, I had something to do with that. Really, that's the only reason why we're here. The other people aren't that important. It's our students that we help make some achievements. (Kentucky middle school, 40-04; sixth grade reading)

Realism and directiveness. While many teachers expected their schools to improve in the near future, optimism was much more muted when teachers rated their chances to improve

according to the criteria of the accountability system. For example, on the survey 50 percent of Kentucky respondents found the system's performance goals "very unrealistic." Yet, the directiveness of the system was very high, and large majorities among survey respondents and interviewees professed to act according to the system's directions and demands. With grave doubts about the system's meaningfulness (with regard to validity, fairness, and realism), teachers said to be primarily moved out of compliance with the state and out of concern for professional prestige. Thus, in an Analysis of Variance that tested relationships between three levels of engagement in school improvement and means in the importance, validity, fairness, realism, and directiveness scales, only goal importance and directiveness show strong and significant differences in means. For work effort, a similar analysis shows moderate, but significant differences in means for directiveness, but not goal importance. Interestingly, the (reportedly) more industrious teachers were even more skeptical about the meaningfulness of the system. I conclude from these findings that the accountability system was overall a poor motivator for teachers in the eleven schools on probation, and its strongest motivators were authoritativeness and stigma producing compliance and status anxiety.

Job Commitment. While authoritativeness and stigma may have pushed people to become more involved and increase work effort, the highly motivated were not necessarily more committed to staying at their school (means on the commitment scale do not differ according to levels of engagement and effort). On many occasions, we were struck when apparent leaders in school improvement disclosed to us in the interviews that they were planning to leave. Besides better career options elsewhere, these teachers often bemoaned intolerable pressures they felt obligated to respond to at their school. But when teachers believed more strongly in the meaningfulness of the accountability system, their job commitment was significantly higher.

That is, more committed teacher groups had significantly higher mean ratings on the system's fairness and realism, (but not validity). Among the various factors tested, the factor showing the strongest mean difference for job commitment was "expectation of improvement." Thus, more committed teachers were also more optimistic about their school's prospect. According to the interviews, optimism was an article of faith for some, others reasoned that as a result of probation their school would receive more attention and new resources, but almost nobody mentioned changes in their own classrooms. Improvements were mostly understood as improvement of others or the organization as a whole.

Overall, the eleven schools, particularly the seven Maryland schools, were beset with problems of teacher turn-over. About half of the respondents, again with teachers in leadership positions over-represented, were not certain about staying or were certain to leave. In fact, in many of the Maryland schools the teacher turn-over rate was about 50% from year to year. For those that planned to stay, relationships at school but also the challenge of school improvement and an optimistic outlook were drawing points, but fewer respondents named probation as a positive influence on the school. Primary reasons for Maryland teachers planning to leave were better career options elsewhere and the feeling that their school was "a sinking ship." In the four Kentucky schools, fatigue from the pressures of probation was the most important reason to leave. Given different job market conditions in the two states, it is conceivable that for the Maryland leavers exit options tempered the pressures of probation; lack of such options may have heightened these pressures for the Kentucky leavers. But it should be stressed that probation for respondents in the eleven schools resulted in mild pressure only, and given the sentiments uncovered here, an increase in pressure advocated by some policy makers frustrated

with the presumed lack of effect of probation may exacerbate an already severe commitment problem in these schools.

Summary. The study found that probation may have a weak influence on individual work motivation and an overall negative effect on teachers' job commitment. Probation in the eleven schools primarily evokes an urgency for improvements through a dynamic of compliance and concern for professional prestige. Pressures are mild, but tangible for the more obligated teachers. Although both the Maryland and Kentucky assessments are complex, they lack meaningfulness for the great majority of educators participating in the study. Teachers eschew the spirit of reward calculus, and hold against the standardized verdict of failure a philosophy of personal connection to children, psychic rewards, and incrementalism. As a result, the impetus of the system for self-examination and reevaluation of one's personal educational responsibility is weak. The mildness of pressure and the promise of new resources seem to temper potentially harmful effects of probation on job commitment, but at the same time commitment is a serious problem for the eleven schools that seems to be exacerbated by probation. With turn-over rates of 30 to 50 percent per year, and highly motivated teacher leaders leaving as likely as lesser motivated teachers, one cannot argue that with probation the "right people are leaving."

Organizational Development

Intuitions

Most accountability systems hold whole schools, rather than individuals accountable for higher performance, and it is therefore through school-wide improvement that individuals overcome the label of probation. How could probation work in this way? The label of probation

throws the school in crisis, but at the same time makes people realize that "we are in this together." Intense dialog, perhaps even conflict around the discrepancy between the current situation of the school and the state's performance demands ensue. Eventually the faculty pulls together around a set of shared expectations that are the basis for a formal structure of internal accountability. All parts of the school are evaluated, planning and more vigilant monitoring make the school more effective, and with determined leadership the school learns new strategies to turn itself around and change instruction.

Theories

Deliberately induced crisis and group accountability are the main motivational levers imbued in probation, but these levers often come together with programmatic and managerial mandates and supports for capacity building. This mix is to shape the organization's social interactions and improvement strategies. Organizational theorists have recognized the work group as an important source of work satisfaction, commitment, and productivity (Tannenbaum, 1970; Katz, 1970; Mohrman, Mohrman & Odden, 1996). In the field of education, studies by Rosenholtz, (1991) and Little & McLaughlin (1993) have shown that teachers increase their commitment to, and involvement in, reform when collegial relationships at school are strong, supportive, and innovative.

Very little is known about how group accountability might work in the context of schools (Hanushek, 1994; Malen, 1999). Effective schools research considers the school the most suitable strategic unit for educational improvement. Some authors believe that group accountability may be a way "to motivate teachers and administrators to enact their jobs in a manner that leads to significantly higher student achievement, *sometimes without a commensurate increase in expenditures* [emphasis added]" (Mohrman, Mohrman & Odden,

1996, p. 54). But once the attainment of rewards or the aversion of penalties is tied to the group, rewards and sanctions operate in "weaker" situations, in Shamir's terminology, as the individual reward expectations are dependent on colleagues' capacity and willingness. As a result, group-generated performance motivation must tap into teachers' more broad-based and diffuse commitment to the organization. Moreover, in the literature on high performance organizations the group is usually understood as the basis for rewards or bonuses rather than as the unit that may have to absorb sanctions and penalties. The response of work units to sanctions may flow from individual and organizational processes that are quite different from those at work in high-performance or high-involvement organizations. Responses to sanctions may be more adequately captured by a line of inquiry that places the failing organization and its crisis in the center.

Induced crisis, as a means to rouse a declining organization to focus on its essential service (Meyer & Zucker, 1989) can motivate an organization to learn (Leithwood & Seashore Louis, 1998). As probation throws schools into crisis, they unfreeze. Old routines and mental models are up for internal debate and conflict may arise. A conflict-driven scenario of organizational learning is narrated by Bennett & Ferlie (1994): "A crisis moves awkward issues up agendas...We are likely to see continuing pressure from pioneers, the formation of special groups that seek to evangelize the rest of the organization, high energy and commitment levels and a period of organizational plasticity" (p.11). In schools on probation, the initiative should move to high-performing and highly motivated, perhaps even maverick, groups of teachers and administrators. At minimum, this process entails dialog about the goals of the accountability system, a collective commitment to shared expectations, and formal structures that undergird internal accountability (Abelman, Elmore, Even, Kenyon, & Marshall, 1999).

External threat and induced crisis, however, are not automatic triggers of learning (Levine, Rubin, & Wolohojian, 1981). According to Staw's threat-rigidity model (Staw, Lance, & Dutton, 1981), two organizational responses to threat are likely. If the group believes in the likelihood of success in meeting the new demand from the environment, increased cohesiveness, support for leadership, but also a tendency to uniformity and centralized control occurs. If the group believes in the likelihood of failure, constricted interaction gives way to leadership instability and dissension. But even organizations successfully responding to new stressful external demands tend to reinforce dominant patterns of operation, rather than learn new things, according to this model.

Findings

Our data confirm the motivational power of the group for work effort and engagement in school improvement activities. We composed from survey items several internally reliable scales that captured respondents' perceptions of their faculty: *skills of colleagues, collegiality, principal control, principal support,* and *burden sharing* (three items that specifically ask about group control and sharing the work load).[3] Across the eleven schools, respondents with reported higher levels of engagement and effort also perceived their faculties to be significantly more collegial and their principal strong, i.e. both more controlling and supportive. This contrasts with the relative irrelevance of the accountability system's meaningfulness for different levels of work motivation across the eleven schools.

A comparison between two "moving" and two "stuck" schools from Maryland (Rosenholtz, 1991) corroborates this pattern. In the two moving schools, levels of engagement and work effort as well as school-wide activity levels (identified by field work) were higher;

incidentally the moving schools also managed to increase their scores on the key state assessment (MSPAP) during the study period, whereas the two stuck schools either stagnated or declined. In the two moving schools, perceptions of colleagues' skills, collegial relationships, and principal support were significantly more positive than in the stuck schools. (Principal control and burden-sharing were insignificant). By contrast, the meaningfulness of the accountability system was actually more in doubt in the two moving schools. Yet, respondents in the moving schools felt more directed by the system and attached more importance to overcoming probation and raising test scores, that is, they responded more strongly to the extrinsic motivators of the system. Thus, probation made the moving schools more moving due to higher internal organizational capacity and external pressure, i.e. stigma and authoritativeness, but not because they believed in the rightfulness of the system. But it was troubling that teachers in the moving schools were not significantly more committed to staying.

What was going on in the two moving schools? Patterns of organizational interaction and types of improvement strategies were in many respects quite similar in the two schools, one an elementary school, the other a middle schools. Both were located in the same district which had few of its schools on probation, organized a fairly efficient central Office of School Improvement, and awarded to all its "reconstitution-eligible" (RE) schools between $150,000 and $250,000 in excess of the regular budget.

Both schools were led by seasoned principals who had survived in their position, but felt nevertheless under enormous pressure.[4] With teachers feeling probation only as mild pressure,

---

[3] See Technical Report for details at www.gseis.ucla.edu/faculty/ mintrop (click on Schools on Probation).

[4] Since many reconstitution-eligible (RE) schools in Maryland (including the seven selected schools) improved only marginally or not at all after identification, punitive transfers of principals were frequent. In four of the seven schools in our selection, the RE designation was accompanied with an immediate change of the principal. Two of the four new principals did not survive their first year after RE designation, one was transferred after his second

organizational accountability rested on the shoulders of the principals. Their main response was to increase control and to assemble a leadership team of assistant administrators, instructional specialists, test coordinators, and school improvement resources teachers that were often hired through reconstitution-designated funds. Externally constrained by districts' and states' programmatic and managerial mandates and supports, the fate of probation, internally, was largely decided by the interplay between the principals' leadership, the skills and commitments that the specialists brought to their task, and a largely compliant, but relatively immobilized, increasingly inexperienced, and uncommitted staff.

School B was the moving middle school in the sample. According to staff comments, the RE designation made their principal into a more vigilant manager, overriding the traditional hands-off style with which administrators and staff had traditionally accommodated each other at the school. The principal abolished all faculty and team meetings and called House teams into his office once a week. During these meetings, faculty members were informed and admonished to comply with the principal's expectations and the strategies adopted by the instructional specialists. The faculty's role was to report on task completion. The principal began to visit classrooms regularly with checklists in hand. On his visits, he emphasized behavior modification strategies that could be monitored easily, such as the daily lesson plan, a fixed surface structure of the lesson, seating arrangements, bulletin board displays, the placement of the district curriculum on the teacher's desk, the page opened to the day's curriculum, etc., all of which

---

year. One school had a new principal every year for the three years of data collection. In three schools, the long-term principals survived the RE designation, but they felt highly uncertain in their tenure. One of them subsequently lost her job and chose early retirement, leaving only two principals who survived RE designation in their assignments. One of those two retained his job against the explicit wish of the state department to remove him and one retired two years after his school's probation designation. By comparison, across the four Kentucky schools, the situation was more stable. Three of the four schools were headed by principals with long tenure in their schools. One school, by

staged the teachers' compliance with school improvement efforts. In addition, the very skillful instructional specialist had compiled a handbook of generic strategies that she believed would "crack" the complexity of the performance-based test, such as a particular surface structure for essays, a particular way of writing out a math problem, etc. During weekly campaigns, teachers were expected to practice these strategies with their students and were monitored on their use. The school raised its test scores substantially for one year, but was unable to keep up this upward trend.

Control at School B came with a smiling face. The principal was warm and paternal, but had made it clear that they, he with the rest of the school, had their backs against the wall. The instructional specialist was accommodating and always full of ideas, but the teachers knew that her proposals were what the principal wanted to get done. Teachers at School B felt controlled and supported. Many of our interviewees empathized with the principal's difficult position (it somehow reflected their own), they "understood" that accountability dictated stronger measures, and they appreciated the sense of direction that was provided for them, but at the same time many wished to escape the pettiness and pressure and work somewhere else. After the first year of probation, and despite the school's success, 70 percent of the teachers were contemplating to leave the school, and turn-over rates remained stubbornly high.

School C, the "moving elementary school" responded similarly to probation. But here the principal relied more on his team for results. One of the instructional specialists managed to develop detailed knowledge of the school's test score data and designed a daily curriculum for all lower-grade teachers. Many appreciated the support, and some were "encouraged" to teach

---

contrast, had a new principal every year in the last six years, though this was not attributed to the school's performance status, but to district problems.

these lessons by the principal's unannounced visits. However, although the school managed to improve as evidenced by raising test scores substantially in two consecutive years and the district alleviated severely overcrowded conditions, teacher turn over hardly abated. Out of 30 classroom teachers in 1997/98, only eight could still be found in the school at the beginning of the 2000/2001 school year; of those eight, four were kindergarten teachers. The principal had announced his retirement and the instructional specialist her leaving.

A third school, a middle school located in the same district, was also moving, but it moved astray into "pathological rigidity" as we termed it. With a legacy of discipline problems, the district installed a new principal with a proven track record of school improvement who brought with her a loyal leadership team and cleaned house. Her hallmark were a tight hall supervision policy and the same control mechanisms in use in the other two schools, minus the attention to curriculum and instruction. Debate was not tolerated at faculty meetings, and teachers' rule infractions were publicly rebuked and justified with the need for accountability, occasionally over the PA system. Test scores never improved much or declined, and year after year fifty percent or more of the faculty left. In the end the leadership team imploded and the remaining teachers threatened a walk-out.

In the two "stuck" schools, probation was an altogether less dramatic affair at the time of the study. Located in a district where support for probation schools had to be spread over half of the district's schools, Schools D and E had been designated for three years at the time of the study and probation had become habituated. If School D, a large middle school, had ever shown a more spirited response to probation, there was no trace of that during the time of our field work. MSPAP test scores had remained very low and stagnant for the entire probation period, and for teachers the signal of probation was simply submerged among the many other concerns

for daily order and survival. When we first entered School D, we encountered a dispirited principal who felt he had barely made a dent in his school during his one year tenure. Frustrated by flat test scores, district inaction on the most basic building repairs, and feuding with the faculty, he was counting his days to be replaced. The next principal showed very little urgency and concern for change. He said that he would study the school the first year and then take his steps. He was liked, but he was also known to take his breaks with other teachers smoking under a tree off school grounds. His tenure ended with an acrimonious faculty meeting during which some faculty members aired their raw frustrations with his inaction.

School E, a small elementary school located in a very poor section of town, had actually been moving at some time. When the school was "named" as one of the first in the state, the school and community had organized a spirited rally in support of the school and in protest against the state. The principal, rooted in the community, had a background in staff development, but management was not her strong suit, so she opted for intensive training of her staff in performance-based pedagogy. Test scores improved remarkably and with increasing numbers of schools entering probation, she was in high demand as a speaker and trainer. But her training-based improvement model faded as the district spread staff development resources thinner with increasing numbers of probation schools, and as high teacher turn-over erased past training gains. The school was unable to fill positions, and it became common that teachers quit mid-year. While on probation, the percentage of inexperienced teachers increased to 70 percent. One year, with test scores plummeting, the school could not fill positions with permanent teachers in 3rd grade, a key testing grade for MSPAP. Also, while on probation, the percentage of special education students increased from 23 percent, already above the district average, to 27 percent. All instructional specialists had to be moved into regular instruction, and as a result the faculty

according to one of the specialists did not even have sufficient basic capacity to implement the district-mandatory and very prescriptive "Open Court" literacy program even with district training. With test scores decreasing and the faculty dispirited and worn-out, the district decided to replace the principal who hitherto had provided a modicum of stability, with a new, but inexperienced principal. The school declined even further, and after six months into the tenure of the new principal, the state announced that the school would be taken over by the state the following year. That year, satisfactory scores on the MSPAP plummeted to zero or near zero.

Summary from the Maryland Cases

Rather than staging crisis and opening channels of inquiry into solutions with broad faculty participation, administrators are *in* crisis and as conduits of accountability tend to mute the voices of outspoken critics who might question the undisputed reality and legitimacy of the accountability system, but whose ardor might also expose the school to honest self-evaluation. Accountability is accepted as a fact; the value and realism of performance goals is not publicly deliberated in most schools. The teachers are willing to rally around their leader as long as they sense tangible progress. Teachers resist crude managerial control, but accept increased control in those schools where it is laced with traditional paternalism and concrete assistance. Teacher learning takes place as skill (re)training primarily.

High principal turnover or low-impact principals doom a school's probationary period. Our "success" cases have higher (perceived) capacity. They are more unified and supportive, their faculty is perceived as more skillful, and capacity building is seen as more effective. But also, probation makes teachers compliant, and traditional prerogatives of teachers' classroom autonomy are overcome through administrative power attached to specialists' instructional support. In the more successful cases, increased rigidity is associated with more effectiveness of

the organization. Discipline tightens; more attention is paid to the state assessments; classroom teachers are on guard. Career teachers and instructional specialists are roused into action and rally around the principal. A curriculum is being followed. Increased participation in staff development workshops may have increased the competence of (especially novice) teachers. Some of the seven schools post modest improvements in this way. But increased organizational rigidity exerts a price. Teachers are dissatisfied; some resent the pressure and standardization, and many contemplate leaving. These are not circumstances under which internal accountability flourishes. Improvement strategies chosen by the schools correspond to the patterns of leadership. Schools rely on external programs, sweeping standardization, easily surveillable behavior, surface compliance reviews, and test preparation schemes (see also Mintrop & MacLellan, 2002), for results of the content analysis of school improvement plans).

Contravening Evidence from the Kentucky Cases

On the individual level, notwithstanding different weights, basic patterns were similar in schools from both states. But on the organizational level, we did not observe the same kind of organizational rigidity pattern in the four Kentucky schools that were so prevalent in the Maryland schools. To begin with, probation in the Kentucky schools was an altogether less stirring affair. (This is 1998 when the first wave of accountability demands is spent and the system is in the throes of political contestation.) Teachers and administrators stressed continuity of their school's efforts to improve regardless of the school's status. The Kentucky respondents attached less importance to higher test scores and less meaning to the accountability system. Performance problems tended to be externalized. One school considered itself the "district dumping ground," another the district "special ed magnet." Although public stigma hurt and

instilled in most teachers a desire to shed the "in decline" label, they reported to a lesser degree than Maryland teachers having exerted more effort as a result of probation.

Although, compared to Maryland responses, Kentucky respondents were less optimistic about their school's prospects of improvement and less certain about their efficacy with their students despite higher levels of work experience, they gave their schools higher marks on capacity. Faculties were seen as more skillful and collegial and principals as more supportive and less controlling. Principals themselves did not feel threatened in their jobs based on test scores. Three of the four principals owed their long tenure to districts that the state accountability system largely by-passed, and district interventions were not as prominent. Hence the urgency that fueled control strategies in the Maryland schools was largely absent. Teachers felt challenged to do a better job at aligning their curriculum with the increasingly prescriptive state core curriculum and pay more attention to test-specific features, such as writing prompts. Because the faculties were fairly stable, there was more evidence of training effects in the interviews. The DE/HSEs provided assistance in assessment-specific features and were seen as helpful in keeping their schools focused on what schools could internally control (Mintrop, MacLellan, & Quintero, 2001). But they did not direct the schools' improvement strategies, rather their formal authority position was absorbed into the traditional hierarchy of the schools. More managerial autonomy of the school fostered entrepreneurialism in attracting new grants and projects. "You name it, we've tried it," as one principal termed it, was the visible badge of the schools' commitment to improvement. Several HSE's bemoaned that this approach left (low) expectations and classroom routines in these schools largely unexamined.

In summary, then, the detected rigidity effects of probation in the Maryland case may be related to a specific constellation of factors: More district control, threatened principals, and

ordinary teachers with low skills, low commitment, and modest work motivation all working within a state accountability system that steers local districts with pedagogically complex outcome demands without providing the tools to reach them. In the four Kentucky schools, we observe a more traditional pattern of school improvement through alignment and add-ons.

Instructional Changes

Intuitions

Given the ambitious performance-based character of the accountability systems studied here, schools, in order to master probation successfully, not only need to compel students to work harder, but also learn differently. Higher work intensity, tighter lesson plans, but also higher order thinking and teamwork are paramount. When teachers have the will to change and faculties have begun to evaluate the shortcomings of their school, raise their own expectations to the high demands of the system, and agree on formal procedures of internal accountability, the conditions are ripe for a restructuring of teaching content and methods.

Theories

The literature on curriculum policy and instructional change shows that what teachers learn from policy depends on a host of factors: their extant practices, their understanding and interpretation of the policy, their own experiences, dispositions and skills, and the support they receive in efforts to change their practices (Cohen, McLaughlin, & Talbert, 1993; Darling-Hammond, 1997; Grant, 1998) found that teachers responded quite differently to the same reform, even when exposed to the same interventions. Spillane & Jennings (1997) show that when districts employed alignment strategies to change instructional practice at the level of classroom discourse, they often achieved superficial task modification, but did not reach more

deeply ingrained task and discourse structures. Teachers often trivialize complex tasks to simpler task demands (Cohen, 1990); Spillane & Jennings, 1997; see also Spillane & Zeuli, 1999; Cohen & Ball, 2001)and they doubt the relevance of ambitious performance standards and institutional demands when incongruent with the perceived needs of their students (Darling-Hammond & Wise, 1985). If teachers learn ambitious pedagogy through "revisiting and reinventing" (Cohen & Ball, 1999), then probation cannot succeed without accountability being connected to personal educational meanings and processes of organizational dialog and learning that facilitate exploring these meanings.

Findings

It is apparent from the previous sections on individual learning and organizational development that probation in the context of the eleven schools provided unfavorable conditions for learning new and ambitious performance-based pedagogy. For many teachers, the state assessments did not provide meaningful tools for the self-evaluation of their teaching. On the organizational level, probation fostered rigidity and compliance with external obligations, as in the case of the Maryland schools, to the detriment of organizational learning and internal dialog. Moreover, whereas the accountability system calls for an upgrading of teaching quality, the investigated schools on probation struggled with high teacher turnover, low job commitment, and an increasing number of uncertified and inexperienced teachers.

We saw that large numbers of teachers in the 11 schools viewed themselves as highly competent professionals whose skills and knowledge measured up to the demands of the states' performance-based assessments. But in reality, 70 to 80 percent of the observed lessons in Maryland did not show evidence of elaborate level teaching at all; that is, the frequency of higher-order thinking, problem solving, and complex dialog among the counted snapshots (total

number 150) was very low. (Due to their limited number, we did not quantify the observed lessons in Kentucky.) Only one third of all observed Maryland lessons were deemed highly coherent, i.e. beginning, middle, and end hung together; the majority lacked conceptual depth. On the positive side, in the overwhelming majority of lessons, teachers used a variety of materials, activities, and forms of interaction; in quite a few lessons variety was a very prevalent feature. Contrary to some other assessment systems (Darling-Hammond, 1991; Noble & Smith, 1994), evidence of practicing simple test taking skills (i.e. "drill and kill") was fairly low.

A selection of seven teachers illustrates patterns observed across the 30 classrooms. We observed the classroom of a senior middle school teacher who had the reputation as an innovator. In the observed lesson, she had students measure the relationship between diameter and circumference of various circular objects. By following the lesson "script" from the newly adopted mathematics textbook, she believed that her instruction was aligned with MSPAP because "they match the skills with the national standards. So it is really close because I know the MSPAP is taken from the state standards that they get from the national standards" (G-16). The problem was that, contrary to the intentions of the book, she herself introduced Pi without the students having had a chance to discover the relationship themselves. She said, "I'm a math teacher. I'm used to, you know, this is this and this is this..." To her, the ability to work successfully in groups was the key to MSPAP proficiency, construction of concepts was the lesser of her concerns. But despite her willingness to change, accountability for her was "treating you more like a child... So, I see it as the work doubling....And we do it, I mean, you know, they say "this" and we do it. [But] the morale doesn't work very well. (G-16)

A young and effective mathematics teacher in another middle school coped with accountability differently. Because of the variety of testing situations she had to prepare her

students for, she parsed her lessons in a regular pattern of basic skills, performance-based, and regular lessons: "If it's just a regular lesson, no MSPAP, no Functional per se, then I rely on the book" (B-17). She was obviously a skilled classroom manager and adept at teaching. When asked how she envisioned closing the gap between her students' capacities and the state's expectations, she responded: "Well, one day at a time, basically. I know that the gap cannot be tightened within a, you know, short period of time… and in time, if instruction is, if you're doing what you're supposed to do, then the scores will come up" (B-17)

We visited the classroom of a very respected science teacher with the reputation of a disciplinarian. He taught a very directive and repetitive lesson. Changing his teaching in response to the accountability system was out of the question for him. He taught the content he believed needed teaching, with the materials of his choice (very old textbooks), in the manner that he saw fit. An alternation of very directed reading with experiential lessons worked best for him. Similarly, his colleague, a middle-aged woman, highly respected by students, taught a very traditional lesson that kept students working hard. She introduced herself: "I don't know if you want this on tape but I have a sticker on my car that says, 'Stop MSPAP, teach basics.'" She subverted external pressures and was outspoken about her conviction that she had better sense than the various distant agencies and actors that tried to tell her what worked with her students:

> No matter what happens, I don't change, and they [the students] depend on it. That's important to them. It's also a part of classroom climate. My expectations don't change for them. They know what to expect. (A-22)

An elementary school teacher, in contrast, was delighted with the instructional materials mandated by the school. She loved the scripted nature of the *Open Court* reading. She was confident that if students could read they would be successful at taking the MSPAP test when the time came, or any other test for that matter.

Mr. C. faced the problems that many beginning teachers face. We observed a frustrating lesson during which he attempted to teach the difference between "action and state of being verbs." He received little instructional support or guidance. Student discipline was not, in his view, an administrative priority at his school and that combined with parental non-involvement made classroom management very difficult for him. He faulted his inability to raise his instruction to grade level on his students' lack of knowledge. Mr. C. assured us that "I can be rather creative when I'm in the right environment." When asked whether the reconstitution eligibility status of the school or the MSPAP influenced his teaching, he replied, "Is MSPAP driving what we're doing in any way? No. No. What's driving what we're doing is survival" (G-21). Another beginning teacher, Mr. S., was in the same situation. Nonetheless, he made an effort to follow the adopted curriculum as closely as possible, but was often thrown for a loop, as in the observed lesson, because the curriculum did not match up with his students' below-grade-level skills. He did not think that students could come up to the expected performance level because his third grade students "already so far behind" (F-20). Mr. S. was, in his words, "worn out" and "run down" by trying to reconcile the reality that "a lot of things that the kids come into school with are things that are way beyond my control" and being held accountable for student achievement. He intended to leave teaching at the end of the school year.

Most teachers we visited considered it unlikely that their students would reach the lofty goals the accountability had set out for them. But teachers were willing to try concentrating on incremental learning steps. In negotiating the gap between external performance demands and the perceived abilities of their students, teachers foremost gauged their lessons to students: "I teach to the needs of the students. That's what I feel I should do because if I try to teach up here and they're not up there, I'm wasting my time and my energy because they will never meet with

success" (F-17; second grade). Teachers felt justified teaching lessons in a basic skills format that traditionally "worked" for "their kids." In the view of many, MSPAP activities distinguished themselves mainly as writing activities, group work, and the use of particular analytic vocabulary. For fewer teachers, reflection on one's own thought process was also associated with MSPAP. This pattern holds across all observed teachers. Thus, often the conceptual depth of knowledge construction that is a core element of the new pedagogy was simplified into a set of activity formats. Judging from the debriefing interviews that accompanied lesson observations, teachers were, for the most part, not aware of this task trivialization. But this was not surprising, considering that the test itself was shrouded in mystery and teachers only reluctantly discussed items that they had seen for fear of doing something inappropriate.

Although teachers strongly expressed the notion that their lessons were first and foremost adapted to their students' ability and achievement levels, tests, new instructional programs, new curricula, and new textbooks reached deeply into many teachers' classrooms. But external pressures and directions were multiple and often contradictory. In the survey, more than half of the teachers had indicated that they were clearly directed by the accountability system. In the context of their classrooms, this clarity appeared to be much more laden with conflict.

For all their resentment, many teachers, almost in passing, expressed habitual compliance with administrative mandates intended to align instruction with MSPAP. Although they saw the accountability system as unhelpful and stacked against them, they did not reject it and did not outright condemn it. They truly served two masters. They wanted to concurrently accept the institutional weight of the state and be sensitive to the needs of their students, but the two pulled from opposite ends. Some teachers learned from this tension, but more frequently tension was diffused by discarding the state's directives by virtue of their unreasonableness, or by discarding

students as uneducable. But the great majority adopted officially sanctioned programs as a defensive retreat that relieved them of dissonance and delegated the decisions and responsibilities to a higher level. In all the schools, a main feature of instructional reform was the monitoring of surveillable behavior. A few teachers considered these "good pressure" (F-18), but many experienced teachers thought they failed to attack the real problems faced by the school, like this senior teacher:.

> We are now held to certain standards and expectations such as we have expectations for the students, the principal has some and administration has expectations for us and she checks to make sure that these things are being done....In the past we've more or less been left to do to our own task and like I used to be guilty of not writing lesson plans. I'd come in, I knew what I wanted to teach....Really, the only thing I do differently now is write out.....But other than that, there's nothing I didn't do that I've changed. I still have my objectives, my outcomes, my warm-ups....That was all there before. (A-24)

Instructional Change in the Four Kentucky Schools

Descriptions of Kentucky classrooms are based on a smaller base of observations. Therefore Kentucky patterns are more anecdotal. None of the Kentucky teachers we visited in their classrooms reported instructional sea changes as a result of the state assessments. Most teachers described their instructional changes as "add[ing] skills here and there" (10-10, second grade math). Patterns of instruction in the small number of classrooms we visited in the Kentucky schools were very similar to the patterns encountered in the Maryland schools. Most lessons were taught on a basic level, some lessons were marginal.

Changes mentioned by teachers bifurcated. There were those that had to do with alignment. Overall Kentucky interviewees seemed relatively well informed about unique features of the system, most notably portfolios and writing prompts. Conceivably, with less staff fluctuation, professional development and assessment-related training may have left their mark

to a much larger degree than in the Maryland schools. Highly Skilled Educators cautiously focused schools on key tasks of the reform, such as planning, portfolios, writing, and curriculum alignment. Most teachers said they tried to cover the core curriculum as best they could:

> Well, each school does an aligned curriculum, so that's what I'm supposed to teach. That's my aspect of it....What we try to do is make sure that we have given them a thorough review for the test. We try to get as much through as possible. With the test, they give you roughly what percentages, like 10 percent is going to be weather and stuff like that, so you say, "Well, okay, it's going to be 10 percent weather, so I can give them worksheets on weather and give them a project on weather." Things like that....(40-15, sixth grade science)

But on the other hand, Kentucky schools had wide discretion in selecting programs and strategies. For some teachers, instructional goals were to raise achievement on state assessments while many others revealed that the need to motivate students was a more persuasive influence on their decision making about instructional materials. Pressures from the school's accountability status did not seem to foreclose their own approaches and curriculum alignment left flexibility. Overall, being "in decline" did not exert a strong press towards broader instructional changes. One teachers weighted the effect of the "in decline" status this way:

> Yes and no. Yes because there are certain things that I must do to... They have requirements now for us, like we have to do so many questions with the kids, like open response questions and things like that, so yes, it affects me in that way. The curriculum is not changed any. The curriculum is the same. If they were not here, if I'm doing my job, then I'd still be doing the same curriculum. Channel 11 was in here and they asked the same question. They said, "What are you doing...blah blah blah?" And I said, "Well, the academics are still the same. It's just our accountability that's up for grabs right now." We have to show that we're doing these things by performing, jumping through certain hoops. But no, I don't think I'm a better teacher because I'm in decline trying harder.....I don't think I am doing anything different. I do post stuff, though, because I'm supposed to. .....I didn't have all the nice little things that they give us, so now I can label everything I do.... We have to succeed on a test that the students take. .....The students will take the test, but their grades are going to have to be high enough that we succeed. (40-16)

Neither in the Maryland nor in the Kentucky classrooms did we encounter much of the mind-numbing test drill and practice that have been reported from accountability systems in which traditional basic skills tests have become high-stakes.

The Limits of Sanctions

Incentives and sanctions are the linchpin of a new generation of high-stakes accountability policies. I have explored how schools in pedagogically complex accountability systems responded to the signal of probation. On the positive side, almost all of the eleven schools were modestly energized by the label, at least at some point or from time to time. Teachers in all schools reported that they increased work effort and engagement in school improvement as a result of pressure and direction. Management in some schools tightened up, educators paid closer attention to the state assessments, and the adoption of new programs, strategies, and projects accelerated. In this way, a number of schools were able to remedy some inefficiencies and provide more structure to teachers than previously had been there.

A look at MSPAP test scores across the seven schools from Maryland during the post-probationary period suggests as much. While two schools have made notable strides on the MSPAP test in the areas of math and reading since becoming "reconstitution-eligible," for the majority MSPAP performance in these key areas has either seen a very modest increase or decline, but mostly stagnancy. Additionally, schools have been plagued with year-to-year score fluctuations. But all seven schools arrested decline in the first years after identification. In all these respects, our sample resembles the overall patterns identified for all Maryland schools on probation in the 1996 and 1998 cohorts. We concluded from this analysis that probation may foremost be a tool to arrest decline in persistently low-performing schools, but it may not produce large gains. By remedying gross inefficiencies, many schools are able to "harvest the

low-hanging fruit," as one of my colleagues calls this stage, but they make few further inroads into the territory of instruction.

Had the assessments been less complex and more basic skills oriented, probation may have "worked" better, that is, the pressure of the stigma combined with various control strategies and program standardization I described earlier could have produced an intensification of instruction based on already existing competencies and instructional formats. But pressure is a double-edged sword. It may challenge people to increase work effort, but also make them want to leave if they do not value the pressure as serving a worthy purpose. Increasing pressure would have exacerbated the already immense problem of job commitment in the studied schools. Moreover, teachers knew that they did not face too much competition for work places that many outsiders viewed as unattractive. Hence, as to sanctions, most teachers called the states' bluff. Principals, however, standing in for the accountability of the organization, felt differently.

Probation was not working well as a tool for instructional reform. To begin with, majorities of teachers did not find the standards of the accountability systems meaningful for their work as educators. Being "unfairly" branded as "deficient" by a system that was seen as insensitive to the needs of their disadvantaged students, they in turn based their own sense of worth on personal relationships in their close-up environments. Rather than accepting criteria and judgments of the system, they felt singled out as the ones who had to carry the "blame" for student learning and in turn externalized the causes for underperformance. Probation did instill in schools the notion that "something" had to be done, but in none of the schools did probation trigger elements of internal accountability, if this is to mean a process through which a faculty formulates its own expectations in light of student needs and high demands of the system, agrees upon formal structures that hold them to account, and focuses improvement on identified key instructional deficiencies.

This kind of internalization process was neglected in the Kentucky schools, but forestalled in the Maryland schools. The rigidity pattern encountered in the latter is an example of what happens when high performance demands and top-down pressure meet low capacity schools. The result was a proliferation of control strategies that had the potential to turn classrooms into the opposite of what performance-based pedagogy intended. Being "in the fish bowl" most teachers tightened up traditional lesson structures. Coverage and task completion reigned supreme and more group work and writing assignments were added. Looked at through the perspective of the seven focal schools, the Maryland case illustrates the limits of steering educational reform through incentives and sanctions placed on outcomes without an instructional technology that facilitated the alignment of demanding outcomes with curricula and materials and provided a bridge to student needs. The state also left capacity building largely in the hands of local districts, with the result that external demands and pressures fell upon wholly unprepared schools that reacted with rigidity, rather than learning. In the Kentucky cases these responses were avoided. Here, by contrast, higher capacity schools responded to a system that was less ambitious pedagogically, more prescriptive as to alignment, and more supportive through the HSE feature.

But there are success cases among the Maryland schools on probation. And one of the moving schools in our sample may give us an idea of what went into their improvement: an experienced principal, exceptional instructional specialists with data analysis, curriculum development, and coaching abilities, and additional resources provided by the district. And yet, prospects for the school are dimmed. Leaders exit, and without a process that involves all faculty members as responsible and committed actors, rather than mere implementers, the school may find itself at a loss again.

The accountability systems in both states operate on the assumption of organizational stability. Only this assumption makes it legitimate to publicly expose putative deficiencies of whole schools based on year-to-year comparisons of schoolwide test scores. The reality of the 11 schools on probation selected for the study is, however, quite different. These schools were, on the whole, hard and challenging work environments that educated large numbers of students considered at risk. Particularly the 7 Maryland schools faced instability due high teacher and administrator turn-over, and increasing proportions of inexperienced teachers which first and foremost raises the specter of student discipline problems and social instability. But the Kentucky schools as well were beset with high student mobility rates and changing student intake or attendance zones, changes that were outside of their control. Under these conditions, continuous improvement was an impossibility due to the schools' lack of organizational continuity. Thus, many of the schools, particularly in Maryland, needed baseline stabilization first before they could embark on ambitious instructional reforms. The responsible actors for this kind of stabilization are for the most part districts and states. Teachers themselves give us an idea of what is urgently needed. When asked to select among 10 priorities for school improvement, large majorities chose student discipline, teacher motivation, and teacher turn-over while not even ten percent believed that a new pedagogy should be first on the agenda.

Note:

I would like to thank Daria Buese, Kim Curtis, Lea Plut-Pregelj, and Margaret Quintero for their assistance. This research was supported by a grant from the Office of Educational Research and Improvement, U.S. Department of Education (No. R308F70035). The views expressed in this paper are mine.

References

Abelman, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When Accountability knocks, will anyone answer?* (CPRE Research Report Series RR-42). Philadelphia: Consortium for Policy Research in Education - CPRE.

Bennett, C., & Ferlie, E. (1994). *Managing crisis and change in health care: The organizational response to HIV/AIDS*. Philadelphia: Open University Press.

Boser, U. (2001, January 11, 2001). Pressure without support. *Education Week, 20*, 68-71.

Chicago Public Schools (1997) *School Probation Information Packet*. Chicago: Office of Accountability.

Cibulka, J., & Lindle, J (2001). *The politics of accountability for school improvement in Kentucky and Maryland: Technical Report, Volume IV* . Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Cohen, D., & Ball, D. (1999). *Instruction, capacity and improvement* . Philadelphia: Consortium for Policy Research in Education (CPRE Research Report).

Cohen, D., & Ball, D. (2001). Making change: Instruction and its improvement. *Phi Delta Kappan, 83*(1), 73 - 77.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation & Policy Analysis, 12*, 311-329.

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan, 73*, 220-225.

Darling-Hammond, L. (1997). School reform at the crossroads: Confronting the central issues of teaching. *Educational Policy, 11*(2), 151-166.

Darling-Hammond, L., & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85*(3), 315-336.

David, J., Kannapel, P., & McDiarmid., C. (2000). *The influence of distinguished educators on school improvement: A study of Kentucky's school intervention program* . Lexington, KY: Partnership for Kentucky Schools.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Fuhrman, S., & Elmore, R. (2001). Holding Schools Accountable: Is It Working? *Phi Delta Kappan, 83*(1), 67-72.

Fuhrman, S., & Odden, A., R. (2001). Introduction: A Kappan Special Section on School Reform. *Phi Delta Kappan, 83*(1), 59-61.

Grant, S. (1998). *Reforming reading, writing, and mathematics: Teachers' reponses and the prospects for systemic reform*. Mawah, NJ: Lawrence Erlbaum.

Hanushek, E. A. (1994). *Making schools work: Improving performance and controlling costs*. Washington, D.C.: Brookings Institution.

Hess, G. A., Jr. (1999a). Expectations, opportunity, capacity, and will: The four essential components of Chicago school reform. *Educational Policy, 13*(4), 494-517.

Hess, G. A., Jr. (2000). *Changes in student achievement in Illinois and Chicago, 1990-2000* (An occasional paper released by the Brown Center on Education Policy). Washington, D.C.: The Brookings Institution.

Katz, D. (1970). The motivational basis of organizational behaviour. In V. Vroom & E. Deci (Eds.), *Management and motivation* (pp. 265-292). Hammondsworth, GB: Penguin.

Kelley, C. (1999). The motivational impact of school-based performance awards. *Journal of Personnel Evaluation in Education, 12*(4), 309-26.

Kelley, C., Conley, S., & Kimball, S. (2000). Payment for results: Effects of the Kentucky and Maryland group-based performance award programs. *Peabody Journal of Education, 75*(4), 159-199.

Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly, 33*, 474-505.

Lawler, E. (1973). *Motivation in work organizations*. San Francisco: Jossey-Bass.

Leithwood, K., & Seashore Louis, K. (Eds.). (1998). *Organizational learning in schools*. Lisse, NL: Swets & Zeitlinger.

Levine, C. H., Rubin, I. S., & Wolohojian, G. G. (1981). *The politics of retrenchment: How local governments manage fiscal stress*. Beverly Hills: Sage.

Little, J. W., & McLaughlin, M. W. (1993). *Teachers' work: individuals, colleagues, and contexts*. New York: Teachers College Press.

Malen, B. (1999). On rewards, punishments, and possibilities: Teacher compensation as an instrument for education reform. *Journal of Personnel Evaluation in Education, 12*(4), 387 - 94.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Meyer, M. W., & Zucker, L. G. (1989). *Permanently failing organizations*. Newbury Park: Sage.

Mintrop, H., & MacLellan, A. (2002). School improvement plans in elementary and middle schools on probation. *Elementary School Journal, 102*(4), 275 -300.

Mintrop, H., MacLellan, A., & Quintero, M. F. (2001). School improvement plans in schools on probation: A comparative content analysis across three accountability systems. *Educational Administration Quarterly, 37*(2), 197 - 218.

Mohrman, A. M., Jr. , Mohrman, S. A., & Odden, A. R. (1996). Aligning teacher compensation with systemic school reform: Skill -based pay and group-based performance rewards. *Educational Evaluation and Policy Analysis, 18*(1), 51-71.

Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review, 67*(1), 41-74.

Noble, A. J., & Smith, M. L. (1994). Old and New Beliefs about Measurement-Driven Reform: "Build It and They Will Come". *Educational Policy, 8*(2), 111-136.

O'Day, J. A., Goertz, M. E., & Floden, R. E. (1995). *Building Capacity for Education Reform* . New Brunswick: Consortium for Policy Research in Education, U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 250-312). San Francisco: Jossey-Bass.

Ofsted. (1997). *From failure to success: How special measures are helping schools improve* . London: Office of Standards in Education.

Pankratz, R. S., & Petrosko, J. M. (Eds.). (2000). *All children can learn: Lessons from the Kentucky reform experience*. San Francisco: Jossey-Bass.

Peterson, P., Rabe, B., & Wong, K. (1991). The maturation of redistributive programs. In A. Odden (Ed.), *Education policy implementation* (pp. 65-80). Albany, NY: Suny Press.

Porter, L., Lawler, E., & Hackman, R. (1975). *Behavior in organizations*. New York: McGraw-Hill.

Quality Counts 2001. (2001). Quality Counts 2001: A Better Balance. *Education Week, 20*(17), 8-193.

Reynolds, D. (1996). Turning round ineffective schools. In D. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.), *Merging Traditions: The Future of research on school effectiveness and school improvement* . London: Cassell.

Rosenholtz, S. (1991). *Teachers' work place: The social organization of schools*. New York: Teachers College Press.

Shamir, B. (1991). Meaning, self, and motivation in organizations. *Organization Studies, 12*(3), 405-424.

Sklar, L., Scheurich, J., & Johnson, F. (2000). *Equity-driven, achievement-focused school districts: A report on systemic school success in four Texas school districts serving diverse student populations* . Austin, TX: Charles A. Dana Center.

Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy. *Teachers College Record, 98*, 449-481.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reform. *Educational Evaluation and Policy Analysis, 21*, 1-27.

Staw, B. M., Lance, E., & Dutton, J. (1981). Threat-rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly, 26*, 501-524.

Tannenbaum, A. (1970). The group in organizations. In V. Vroom & E. Deci (Eds.), *Management and motivation* (pp. 218-225). Harmondsworth, GB: Penguin.

Vroom, V. H. (1964). *Work and motivation*. New York: John Wiley & Sons, Inc.

Wilcox, B., & Gray, J. (1996). *Inspecting Schools: Holding schools to account and helping schools to improve*. Buckingham: Open University Press.

Wong, K., Anagnostopoulos, D., Rutledge, S., Lynn, L., & Dreeben, R. (1999). *Implementation of an educational accountability agenda: Integrated governance in the Chicago public schools enters its fourth year* . Chicago: University of Chicago. Department of Education.

Wong, K. K., Anagnostopoulos, D., & Rutledge, S. (1998). *Accomodation and conflict: The implementation of Chicago's probation and reconstitution policies* (Paper presented at the 1998 Annual Meeting of the Association for Public Policy Analysis and Management). New York: Association for Public Policy Analysis and Management.